# Multi-level Multi-scale Deep Feature Encoding for Chronological Age Estimation from OPG Images

Sultan Alkaabi and Salman Yussof
Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Malaysia
Email: PT20740@utn.edu.my

***Abstract*—Age estimation is a complex task in forensic dentistry especially if the bodies have started to decompose. However, when the task involves Manually examining, the accuracy can decrease due varying experience of the experts, the results of different experts may also vary. To improve speed and accuracy of the age estimation process using forensic dentistry, researchers have proposed Convolutional Neural Network for Dental Age and Sex Network estimation (DASNET). However, pooling and scalar outputs of CNNs could not allow to get the equivariance due to the dental extraction complexity from panoramic images including jaws, teeth, lesions and carries. So, a deep auto-encoder decoder architecture has been developed by the authors, which estimates the age based on semantic and structural feature representation. The age ranges are chosen based on the structural variation of the jaw in these particular age ranges as compared to each other. The authors have proposed a Convolution Long Short Term Memory (ConvLSTM) to capture the correlation of features and generates high level representation of features. For the representation of the generated features, authors have utilized "Atrous pyramid convolution" to produce a multiscale representation. The authors have proposed a combination of multi-scale and multi-level architecture for age estimation. First comes the first sub-part of the model that is the multi-level architecture, it is used for the extraction of hidden features. After that, the output is fed to second subpart which is the multi-scale architecture that enriches the model representation capability in encoding structural and shape characteristics. The propose techniques successfully reduces mean error to 0.75 years, as opposed to 0.93 years of the DASNET.**

***Index Terms*—age estimation, forensic dentistry, DASNET, CNNs, ConvLSTM, Atrous pyramid convolution**

## I. Introduction

Age estimation of human beings in legal medicine and forensic anthropology is a very active research topic [1]. While clinical interest is biggest in youngsters near pubescence, e.g. to survey endocrinological infections [2] or to design muscular mediations [3], [4], interest from legal medicine is an expansive age range around the majority age, i.e., somewhere in the range of 13 and 25 years. As of late, majority age classification of youngsters and youths without legitimate recognizable proof reports relocating to the European Union has seen a ton of consideration, since it is a lawfully significant inquiry to recognize adult asylum seekers from teenagers. To gauge obscure Chronological Age (CA) in youngsters and teenagers can be researched by non-obtrusive, imaging based radiological strategies, transcendently in skeletal [5], [6] and dental structures [7]. This permits specialists in forensic radiology and dentistry to analyze biological improvement identified with hardening of bones [8] and mineralization of wisdom teeth [9]. Notwithstanding, chronological age estimation is inclined to vulnerabilities [10]. Initially, assessing CA based on the evaluation of biological improvement is naturally restricted because of biological differences among subjects of a similar chronological age [11]. This difference characterizes the least blunder that any technique for forensic age assessment can make. Furthermore, because of visual assessment, accepted radiological techniques for evaluating biological development include intra-and inter-rater fluctuations [12], which can be disposed of by using programming based automatic age assessment.

This has been verified with several test results, discussed in the result section. Our method focuses on the limitations of the most recent DANet model with the aim to improve it, and thus provides the following contributions:

1. The authors have applied a bi-directional ConvLSTM (Convolutional Long Short-Term Memory) inside the network to increase the feature re-usability and to include the multi-level representations.

2. The authors have applied the Atrous convolution on the encoded features space to include local and global information in the final representation. Thus, it helps to incorporate multi-scale-representations.

This paper is divided into six sections. Existing research work related to this paper is summarized in Section II. The proposed techniques on Multi-level Feature Combination and Multi-scale Representation are explained in Section III. The experiment conducted is described in Section IV, while Section V explains the performance parameters that will be used for performance evaluation. Section VI presents the experimental results and provides a discussion of the results. Finally, Section VI concludes the paper.

Manuscript received March 18, 2022; revised July 12, 2022.


doi: 10.18178/joig.10.4.151-157

## II. Related Work

Tekin *et al.* [13], Alijani *et al.* [14], and Gharabagi *et al.* [15] have studied age estimation with 'Hand-wrist,' a manual tool for estimating age. The surfaces of the cranium, pubic bones, and rib ends hold clues. At the microscopic level, investigators can see the bone "remodeling" that takes place throughout life, as well as age-related bone breakdown. Spampinato *et al.* [16] and Liang *et al.* [17] have proposed two automated age prediction methods. Based on the bones of the oral cavity, The crown of a tooth forms first, followed by the root. The age is estimated by comparing the stage of tooth formation in the X-rays and bone with known dental growth standard**s.** Alkaabi [18] presented a methodology for chronological age estimation, which applies a machine learning technique to estimate age based on tooth development. The evaluations use CNN for end to end to address drawback of automated age estimation in forensic dentistry without any transformations. The custom dataset of more than 2000 X-ray images divided to 7 different classes is used for training the CNN architectures. The concept of transfer learning is also used for training the popular CNN architectures like AlexNet, VGGNet and ResNet for age estimation. The performance of age estimation is evaluated by analyzing its recall, precision, F1-score accuracies and average accuracies. Haavikko [19] contributed to dental age estimation by revisiting Gleiser and Hunt [20] to find a link between the alveolar and clinical tooth eruption. Demirjian *et al.* [21] proposed a novel method to identify a set of dental development stages, which can be further used to estimate chronological age. Cameriere *et al.* [22] proposed a modern technique considering the relationship between multi-rooted teeth and regression. Not limited to these approaches, recently Vila-Blanco et. al. [23] proposed a deep neural network for chronological age estimation. The proposed method utilizes the deep structure with regression objective to estimate the age from OPT images (Fig. 1).

Figure 1. Example of an OPG image from a 14-year-old boy.

Considering the auto tagging of music is done in almost a similar fashion to image classification, [24] implemented CNN for feature aggregation using a multi-level fashion since the tagging is highly complex. Firstly they used CNNs of different input sizes to capture the features of the audios in a method called feature learning. They later on, in each layer of the CNN, extract the audio features and sum them up when given a long audio clip. In a Fully Connected Network (FCN) they finally make their predictions of the tags. Their method of multi-scaling features proved successful on datasets with lots of audios. Vila-Blanco [25], our benchmark model, implemented a CNN architecture DANET (Dental Age Network) to try and estimate age and sex. The architecture included utilizing a stack of 2D convolution layers, RELU (Rectified Linear Activation Function) as the activation function and implemented Batch Normalization to scale and adjust the activations. Convolution is a mathematical operation that was adopted in image processing and has proven great success. Each image is represented by a set of pixels. The process of convolution refers to aggregation of each element (pixel) of an image to its neighbors. The concept led to the discovery of Convolution Neural Networks (CNN) which have been used widely for image classification and image feature extraction. By implementing a multi -layer perceptron, the benchmark model contained 3,456,065 parameters. AS an extension to DANET, the researchers developed DASNet (Dental Age Sex Network) DANET to accommodate sex classification. A convolution layer was applied to accommodate the sex and age details.

One of the earliest applications of CNN by LeCun is maybe the LeNet-5 network described by [26] for OCR activities. Their architecture was simple because of the limitation of computing power to train large and deep networks. With the modern increase of computing resources and computing power with GPUs and Tensor Processing Unit (TPU), this has encouraged the development of deep CNN architectures. This is evident by every year people try to develop deep CNNs for image classification. Some of the architectures are RESNET, VGG16, ImageNet among others.

Active Appearance Model (AAM) was one of the first models to be used for the age estimation problem. It was able to extract the features of the faces together with the faces [27]. An Adaptive Boosting algorithm together with Local Binary patterns were adopted in [28] that an ordinal discriminative feature learning was proposed for age estimation. Dong YI [29] developed a CNN architecture for the age estimation problem using a multiscale analysis strategy. When comparing it to the bio-inspired features (BIF), his method contains a deep structure.

## III. Proposed Approach

In this paper, authors have proposed a deep multi-level and multi-scale feature representation approach for chronological age estimation. The model is built on the top of the recent age estimation model to improve its performance. The age ranges are chosen based on the structural variation of the jaw in these particular age ranges as compared to each other. It is primarily focused on the age estimation task as it is highly challenging and applicable in numerous applications in the forensic and medical domain. To motivate our multi-scale multi-level approach, first the DANet structure is presented and then its limitation in age estimation is highlighted. Then the solution is presented to overcome these limitations and improve performance. A layer is a building block in deep

learning that receives input and transforms it to an output using nonlinear functions. The model implements the multi-level and multi-scale architecture. The model uses the DANet Dental Age Network). DANet is a neural network architecture build by the Tensorflow framework. First the image is passed to the conv layer where convolutions are applied for feature extraction, At the top of conv layer, Rectified Linear Unit (ReLU) is applied. After that max pooling is applied on the output. After that the dense connection transforms the activation maps into a single vector. The model accuracy is calculated using the Mean Absolute Error. In-order to learn the hidden features the authors have used a combination of Low-level, Mid-level and High-level representation. The transition layer from each of the encoding block passes the output to the adaptive max-pooling layer, this layer then passes the output to ConvLSTM layer which combines the activation maps and generates a single vector. After the multi-level layer model, comprises of the multi-scale layer which is used for shape and global structural representation, the author has used multi-scale pyramid representation where different conv sizes with different dilations are used, combining all these representations enriches the model representation capability in encoding structural and shape characteristics. The model has been explained in the flow chart (Fig. 2).

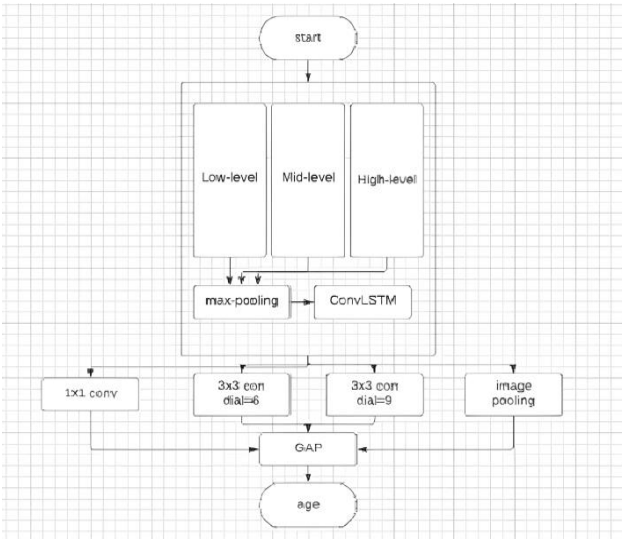


Figure 2. Flow-chart of the model.

The description of the model has been divided into three categories baseline model, multi-level feature combination and multi-scale representation. The categories are explained below.

### A. *Baseline Model*

The Network is comprised of two 3×3 convolution layer blocks where Rectified Linear Unit activation has been applied to them. The convolution blocks used was a substitution of a 5×5 convolution layer, since the output is the same and its performance is almost similar. The model then implements max pooling to encode the input image into feature representation space. On the top of these representations, two convolutional blocks with GAP (Global Average Pooling layer) layers are included to further compress the feature representation into a single vector. Finally, the encoded feature vector is transformed into a single value using a dense connection. The MAE (Mean Absolute Error) between real age and estimated age is utilized to learn the model parameters.

To further boost the performance, the authors train the parallel version of the model (DASNet) where the main model learns to estimate the age and the auxiliary model tries to classify the gender information. Both models use the shared weights transition layers which are included to enhance the feature reusability. Even though the proposed model works well on age estimation, its performance on some particular age range, for example, the age range 15-30 requires further improvement. This is mostly due to the high structural similarity of the jaw are in these pages. Hence, the mechanism to pay more attention to such discriminative information is mandatory.

### B. *Multi-level Feature Combination*

Applying a series of the convolutional block on the input sample can produce a high-level representation of features. Usually, deep CNN models use his high-level representation to learn the hidden patterns. In this work, authors have proposed to use a combination of low-level, mid-level, and high-level representation to increase feature usability. It is proven that combining multi-level representation in a nonlinear fashion can give the model more freedom to learn complex patterns. Thus, in this work authors have included multi-level representation inside the network to increase model capacity in learning-rich information. Fig. 3 shows the structure of the multi-level combination. However, combination the multi-level features in training process takes a lot of time if the computer doesn't have a good GPU. The transition layers from each of the encoding blocks pass to the adaptive pooling layers, where the adaptive pooling layer applies the relative max-pooling kernel to generate a coherent spatial dimension. Once these representations are generated, a ConvLSTM layer combines all these representations to generate the final representation in a non-linear fashion.

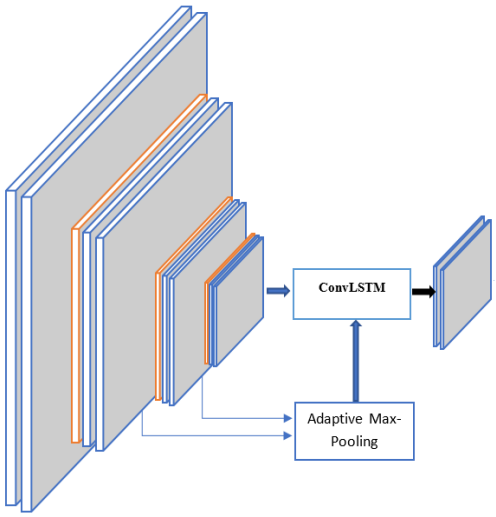


Figure 3. Multi-level representation using ConvLSTM.

A ConvLSTM is based on the Long Short Term Memory (LSTM). The authors have applied the following equations to our model.

$$\begin{aligned} i_t &= \sigma\left(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i\right) \\ f_t &= \sigma\left(W_{xf}X_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f\right) \\ o_t &= \sigma\left(W_{xt}X_t + W_{ho}h_{t-1} + W_{co} \circ c_{t-1} + b_o\right) \\ C_t &= f_t \circ c_{t-1} + i_t \circ \tanh\left(W_{Xt}X_t + W_{ht}h_{t-1} + b_c\right) \\ C_t &= \sigma\left(f_t \circ C_{t-1} + i_t \circ C_t\right) \\ h_t &= o_t \circ \tanh\left(C_t\right) \end{aligned} \quad (1)$$

In equation (1), the input, forget and output gates on time $t$ are determined by $i_t$,$f_t$, $o_t$ respectively. The hidden gate represented by $h_{t-1}$, contains the output of block on time $t-1$ and $C_t$ shows the cell state. The learning parameters of the LSTM model is determined by weight matrix $W$ and bias value $b$ for each gate. And the $\sigma$ stands for the sigmoid function and $\circ$ uses for Hadamard operation. Given the fact that in our proposed method the generated low, mid and high level representations are 3D tensor, it requires modification on the input shape of the LSTM model. Thus, authors have used convolutional LSTM which is proposed for 3D tensor whose first dimension is for channels and the last two dimensions are for the spatial dimension. The reformulation is as equation (2) where the operation applies on the matrix input $X_t$.

$$\begin{aligned} t_t &= \sigma\left(W_{xi} \circ X_t + W_{hi} \circ H_{t-1} + W_{ci} \circ C_{t-1} + b_i\right) \\ f_t &= \sigma\left(W_{xf} \circ X_t + W_{hf} \circ H_{t-1} + W_{cf} \circ C_{t-1} + b_f\right) \\ o_t &= \sigma\left(W_{xo} \circ X_t + W_{ho} \circ H_{t-1} + W_{co} \circ C_{t-1} + b_o\right) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh\left(W_{xt} \circ X_t + W_{hc} \circ H_{t-1} + b_c\right) \\ H_{t=} o_t &\circ \tanh\left(C_t\right) \end{aligned} \quad (2)$$

### C. Multi-scale Representation

Even though the CNN structure is capable of learning local and texture information, it misses the shape and global structural information, which are essential for the jaw area description. The shape information can strongly describe the tooth structure and geometrical information, which are key information for age estimation. In order to include the shape information inside the network, the authors have used the multi-scale pyramid representation. The feature pyramid encodes the feature representation in a different scale, where each scale represents different locality information. Our motivation is that combining all these representations can enrich the model representation capability in encoding structural and shape characteristics. Fig. 4 shows the multi-scale representation structure used in this paper.

This methodology permits the production of dense output feature maps, by maintaining the spatial resolution of each map along the network. This approach results to better performance in segmentation tasks. The implementation of Atrous convolution with distinctive up sampling rates usually overcomes the issue of segmenting objects at multiple scales. Authors have utilized a 3×3 kernel to perform the convolution. A 3×3 mask with a dilation rate, d=2 would have the same field of view as a 5×5 mask as shown in Fig. 4. The reason as to why authors have picked a 3×3 kernel can be analyzed mathematically as follows; Assume you have a square matrix, the kernel, of the order n, where each entry is enumerated with an increasing value from 0 to n2-1. If square matrix size is k= 3, equation 3 below generalizes all of the 3×3 matrices that can be obtained by using a dilation rate $d \in \mathbb{N}$.

$$N_d = \begin{pmatrix} \frac{n^2-1}{2} - d(1+n) \circ \frac{n^2-1}{2} - dn \circ \frac{n^2-1}{2} - d(1-n) \\ \frac{n^2-1}{2} - d \circ \frac{n^2-1}{2} \circ \frac{n^2-1}{2} + d \\ \frac{n^2-1}{2} - d(1-n) \circ \frac{n^2-1}{2} + dn \circ \frac{n^2-1}{2} + d(1+n) \end{pmatrix} \quad (3)$$

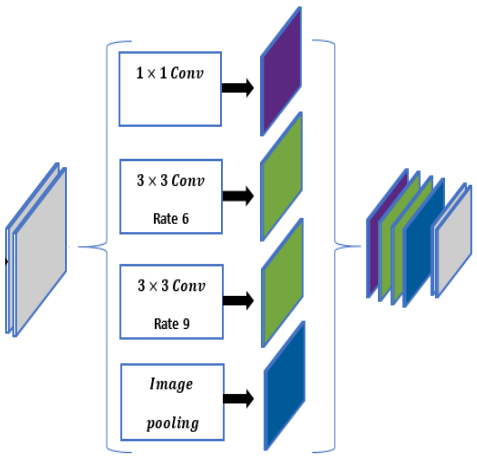


Figure 4. Multi-scale to encode local and global representation.

In our case where the kernel size is $k$=3, and our dilation rate $d$=2, the fields of view order can be found by the equation:

$$k_{ed} = k + (k-1)(d-1)$$

Applying the variables to the equation, our kernel was able to cover a field of view of $k_e$=5.

The dilation rate controls the receptive field size. Therefore, a high dilation rate increases the receptive filed size without requiring a lager kernel size. This characteristic results in generating global representation with the same computation complexity. In this paper, authors have deployed a pyramid of Atrous convolution with four scales. Where the first 3 scales related to local to global representation and the last scale describes the global information of the sample. The pyramid representations are then fed into a point-wise convolution to combine all representation in a non-linear fashion and generates the final representation. Our proposed model incorporates a Global Average Pooling (GAP) layer, which intends to precisely locate the segmented object in the final output feature map by getting the average of each feature map. Lastly, it contains a feature pooling layer.

### D. Model Setup

The overall structure of the proposed method is shown in Fig. 5. The proposed method uses the 4-block encoder

for feature encoding. Then it applies multi-level and multi-scale representation to enrich the feature representation. The generated features are then fed to the Global Average Pooling (GAP) and dense connection layers to produce the estimated age value. Authors have trained the model with parameter $\theta$ using batch size 4 and Adam optimization with learning rate 10e-4. The model optimizes its parameters in each iteration to reduce the MSE loss between real age $a$ and the estimated age $a'$.

$$Loss(\theta) = \sqrt{\Sigma_{i=1}^{n}\left(a_i^2 - a_i^{f\,2}\right)} \quad (4)$$

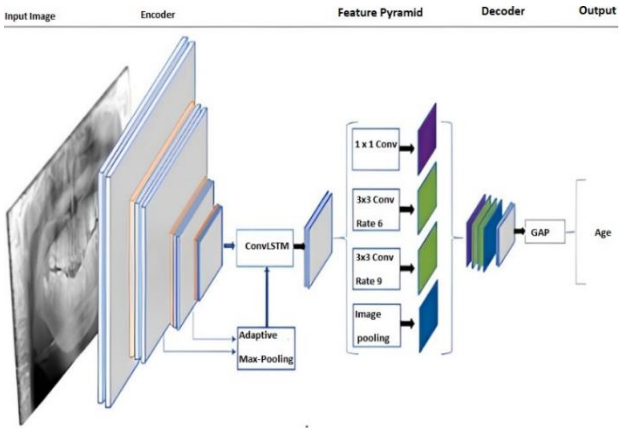


Figure 5. Overall structure of the proposed method for chronological age estimation from OPT images.

With the methods applied, a model is built that has 10,612,866 parameters.

## IV. Evaluation

### A. Data

For the experimental results, a set of 2676 OPG images from Crescent dental maxillofacial and orthodontic services in Bangladesh are provided. The images were labeled with the subject's date of birth and the date the image was acquired, which means that the chronological age was calculated in days. To capture these images, the Osstem implant CBCT T1 device has been utilized. In our experiment, authors have used 70% of the dataset as the training set and the rest as test set.

### B. Experimental Metrics

The performance evaluation is done using the following metrics:

- Age error rate calculated as an absolute error between the real age and the estimated age (AE) - authors measure the magnitude of the measured value subtracted from the true value. The formula can be represented as:

$$\delta = \left| \frac{V_A - V_E}{V_p} \right|$$

where:
$V_A$ = actual value observed
$V_E$ = expected value

- Error distribution information mean ($\mu$) - authors have measured the average error distribution.
- Standard deviation ($\sigma$) - This is a measure of dispersion that measures how data is spread out from the mean.
- 99th percentile ( $p_{99}$ ) - This illustrates the commonality of a result in a comparison to others.

$$R = \frac{99}{100(N+1)}$$

## V. Results and Discussion

In order to achieve comprehensive experiments, authors have considered several experimental settings. In the first setting, authors have applied the proposed method and the state of the art deep models on the main dataset ($S_{all}$) to report the comparative results. Then the dataset is divide into three subsets using particular age ranges, $2 < S_1 < 15$, $15 < S_2 < 30$ and $30 < S_3$. After that, model results are evaluated for each range separately to get an overview of the metrics per category. Authors have chosen these age range based on the structural variation of the jaw in these particular age ranges as compared to each other. Table I shows the experimental results.

TABLE I. Performance Comparison for Chronological Age Estimation

| Model | Dataset | IMG | Med | $\mu \pm \sigma$ | IQR | $p_{99}$ |
|---|---|---|---|---|---|---|
| **Proposed model** | $S_1$ | 264 | 0.41 | 0.53+2.87 | 1.32 | 6.4 |
| | $S_2$ | 759 | 0.45 | 0.67+3.23 | 2.65 | 8.6 |
| | $S_3$ | 1653 | 0.53 | 0.76+4.41 | 2.73 | 14.3 |
| | $S_{all}$ | 2676 | 0.55 | 0.75+4.77 | 3.57 | 13.7 |
| DANet [11] | $S_{all}$ | 2676 | 0.82 | 1.13+5.91 | 4.67 | 16.4 |
| DASNet [11] | $S_{all}$ | 2676 | 0.65 | 0.93+4.85 | 4.01 | 14.6 |

As it is clear from Table I, the proposed method outperformed both DANet and DASNet methods in chronological age estimation. The proposed method estimates the ages with a mean of 0.75 years and a standard deviation of 4.77. Besides that, the proposed method produces a stable estimate for all age ranges. This fact clarifies the effectiveness of the proposed method for smooth age estimation regardless of the age ranges. It is worthwhile to mention that for age range S2, the DASNet generates a mean and standard deviation error of $0.98 \pm 4.83$ while the proposed method estimates ages with mean and standard deviation error of $0.67 \pm 2.23$. Fig. 6 demonstrates the age estimation error in each age range.

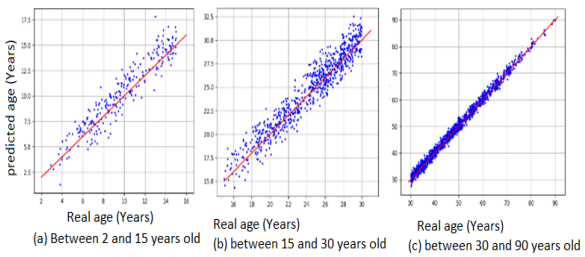


Figure 6. Age estimation error of the proposed method in each age category. The proposed method performs well on each age range.

In Table I authors have presented the performance of the proposed method which shows that incorporating multi-level and multi-scale representation inside the network can increase model performance. Compared to DANet and DASNet, the proposed method achieved a lower rate of $0.75 \pm 4.77$. In order to analyze the effect of multi-level and multi-scale layers separately, authors have performed the experiments on the $S_{all}$ using either one of these layers in combination with the DANet model. Firstly, authors have observed that using DANet+multi-level representation can produce error rate of $0.92 \pm 4.67$ which slightly improves the DANet performance. The main reason for this slight improvement is the fact that mutli-level representation improves the feature reusability in the model. In our second setting, it is observed that using DANet+multi-scale representations can increase the model performance by producing error rate of $0.79 \pm 4.93$. Based on the experimental results, it can be concluded that combining both multi-level and multi-scale representation with DANet can outperform both DANet and DASNet.

## VI. Conclusion

In this paper, an automated deep neural network has been proposed for chronological age estimation using OPT images. The proposed method utilizes both multi-level and multi-scale representation for robust age estimation. Experimental results demonstrated a significant improvement compared to the state of the art approaches.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Sultan Alkaabi designed, implemented the experiments, and wrote the paper. Salman Yussof revised and conducted the research; all authors had approved the final version.

## References


[1] K. Latham, E. Bartelink, and M. Finnegan, *New Perspectives in Forensic Human Skeletal Identification*, Amsterdam, The Netherlands: Elsevier, 2018.

[2] D. D. Martin, *et al.*, "The use of bone age in clinical practice - Part 2," *Hormone Res. Paediatrics*, vol. 76, no. 1, pp. 10-16, 2011.

[3] S. C. Lee, J. S. Shim, S. W. Seo, K. S. Lim, and K. R. Ko, "The accuracy of current methods in determining the timing of epiphysiodesis," *Bone Joint J.*, vol. 95-B, no. 7, pp. 993-1000, 2013.

[4] W. W. J. Wang, *et al.*, "Correlation of Risser sign radiographs of hand and wrist with the histological grade of iliac crest apophysis in girls with adolescent idiopathic scoliosis," *Spine*, vol. 34, no. 17, pp. 1849-1854, 2009.

[5] W. W. Greulich and S. I. Pyle, *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, Stanford, CA, USA: Stanford Univ. Press, 1959.

[6] J. M. Tanner, M. J. R. Healy, N. Cameroon, and H. Goldstein, *Assessment of Skeletal Maturity and Predicion of Adult Height (TW2 Method)*, New York, NY, USA: Academic, 1983.

[7] A. Demirjian, H. Goldstein, and J. M. Tanner, "A new system of dental age assessment," *Hum. Biol.*, vol. 45, no. 2, pp. 211-227, 1973.

[8] U. Baumann, R. Schulz, W. Reisinger, W. Heinecke, A. Schmeling, and S. Schmidt, "Reference study on the time frame for ossification of the distal radius and ulnar epiphyses on the hand radiograph," *Forensic Sci. Int.*, vol. 191, no. 1-3, pp. 15-18, 2009.

[9] H. M. Liversidge, "The assessment and interpretation of Demirjian Goldstein and Tanner's dental maturity," *Ann. Hum. Biol.*, vol. 39, no. 5, pp. 412-431, 2012.

[10] N. Cameron, "Can maturity indicators be used to estimate chronological age in children," *Ann. Hum. Biol.*, vol. 42, no. 4, pp. 300-305, 2015.

[11] T. J. Cole, "The evidential value of developmental age imaging for assessing age of majority," *Ann. Hum. Biol.*, vol. 42, no. 4, pp. 379-388, 2015.

[12] P. Kaplowitz, S. Srinivasan, J. He, R. McCarter, M. R. Hayeri, and R. Sze, "Comparison of bone age readings by pediatric endocrinologists and pediatric radiologists using two bone age atlases," *Pediatric Radiol.*, vol. 41, no. 6, pp. 690-693, 2011.

[13] A. Tekın and K. C. Aydın, "Comparative determination of skeletal maturity by hand-wrist radiograph, cephalometric radiograph and cone beam computed tomography," *Oral Radiology*, vol. 36, no. 4, pp. 327-336, 2020.

[14] S. Alijani, N. Farhadian, B. Alafchi, and M. Najafi, "Relationship of frontal sinus size and maturation of cervical vertebrae for assessment of skeletal maturity," *Frontiers in Dentistry*, vol. 17, no. 20, 2020.

[15] S. Gharabaghi and T. Wischgoll, "A semi-automated method for measuring fels indicators for skeletal maturity assessment in children," *Electronic Imaging*, no. 1, 2018.

[16] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," *Med. Image Anal.*, vol. 36, pp. 41-51, Feb. 2017.

[17] B. Liang, *et al.*, "A deep automated skeletal bone age assessment model via region-based convolutional neural network," *Future Gener. Comput. Syst.*, vol. 98, pp. 54-59, Sep. 2019.

[18] S. Alkaabi, S. Yussof, and S. Al-Mulla, "Evaluation of convolutional neural network based on dental images for age estimation," in *Proc. International Conference on Electrical and Computing Technologies and Applications*, 2019, pp. 1-5.

[19] K. Haavikko, "The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study," *Suom Hammaslaak Toim*, vol. 66, no. 3, p. 103, 1970.

[20] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal dis-criminative features for age estimation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2570-2577.

[21] I. Gleiser and E. E. Hunt, "The permanent mandibular first molar: Its calcification, eruption and decay," *Amer. J. Phys. Anthropol.*, vol. 13, no. 2, pp. 253-283, 1955.

[22] A. Demirjian and H. Goldstein, "New systems for dental maturity based on seven and four teeth," *Ann. Hum. Biol.*, vol. 3, no. 5, pp. 411-421, Jan. 1976

[23] R. Cameriere, L. Ferrante, and M. Cingolani, "Age estimation in children by measurement of open apices in teeth," *Int. J. Legal Med.*, vol. 120, no. 1, pp. 49-52, Jan. 2006.

[24] J. Lee and J. Nam, "Multi-level and multi-scale feature aggregation using pretrained convolutional neural networks for music auto-tagging," *IEEE Signal Processing Letters*, vol. 24, no. 8, 2017.

[25] N. Vila-Blanco, M. J. Carreira, P. Varas-Quintana, C. Balsa-Castro, and I. Tomás, "Deep neural networks for chronological age estimation from OPG images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2374-2384, July 2020.

[26] Y. LeCun, *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541-551, 1989.

[27] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward auto-matic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442-455, 2002.

[28] Z. Yang and H. Ai, "Demographic classification with localbinary patterns," in *Proc. International Conference on Biometrics*, 2007, pp. 464-473.

[29] D. Yi, Z. Lei, and S. Z. Li, "Age estimation by multi-scale convolutional network," in *Proc. Asian Conference on Computer Vision*, 2015, pp. 144-158.

**Sutan Alkaabi** received the B.S. and M.S. degrees in computer Science and information management from the University of Bedfordshire and he started his PHD in data science from University Tenaga National, Malaysia since 2018 and about to finish later of 2022. His research is about hybrid deep learning approaches in medical images. He works as a data scientist in Petroleum Development Oman (PDO) which the leading oil & gas exploration & production in Oman. Mavhine learning and advance technologies are the daily activities with many projects in computer vision, time series, NLP.

**Salman Yussof** received the B.S. and M.S. degrees in electrical and computer engineering from the University of Carnegie Mellon, USA, in 1999, and the Ph.D. degree in engineering from University Tenaga National, Malaysia, in 2010. From 1998 to 1999, he was a Research Programmer with the Institute for Complex Engineered System, Carnegie Mellon University (CMU), USA. Since 1999, he has been a Lecturer with the System and Networking Department, UNITEN University. He is currently an Associate Professor with UNITEN University. He is the author of more than 72 publications. His research interests include computer networks, network security, distributed system, image processing, robotic, and evolutionary computing. In recent years, he has been involved with many professional bodies, such as the Association for Computing Machinery (ACM) since 2003, the Malaysian Invention and Design Society (MINDS) since 2004, the Boards of Engineers Malaysia (BEM) since 2005, the Malaysian National Computer Confederation (MNCC) since 2008, and the Internet Society (ISOC) since 2011.